

# Weekly Report

Lu Junhua

2016 年 11 月 7 日

## Done

- Spent much time overcome jet lag. Very low efficient these days.
- Finished reading the paper to be reviewed.
- Began collecting papers and materials of netease project. The student from Prof. Yong Liu talked about some details about their data. Their data include each action of each player, with timestamps, which we do not posses.
- Wrote the blog of group seminar.

## To do

- Write the review.
- Go to Yvquan to learn more details about the data.
- Collect knowledge on internet/game addiction, mental health (mental disorder) related analysis (using machine learning methods).
- Begin revising the jov paper.

## 论文

Computer Graphics Forum 16 *Faceted Views of Varying Emphasis (FaVVEs): a framework for visualising multi-perspective small multiples*: 对于small multiple提出了一些新的观点和框架, 用实验来验证. small multiple用于编码多维度数据(时间, 空间, 描述性)的时候, 把各种编码叠加起来做比较会比单独并置做比较有更好的效果. 此外提出的针对small multiple的三层抽象模型, 希望不同维度能独立从不同抽象程度来进行更好的比较.

VAST 16 *Visual Analysis and Coding of Data-Rich User Behavior* 是德国斯图加特大学Daniel Weiskopf那边的今年vast会议论文. 文章介绍的是一个对用户行为进行分析(这里叫coding, 用概括性的表征性的短语描述某个行为和动作) 的系统. 完成了下述的任务: 不同层级的coding及展示; 可视化增强的不同角度的探索多源的数据; 不同code之间比较. 文章本身比较完整, 设计也还不错(可能只是不够给人一种完全新颖的感觉? 所以才算会议论文), 我比较喜欢那个word-sized 的对活动编码的设计, 如下图1.

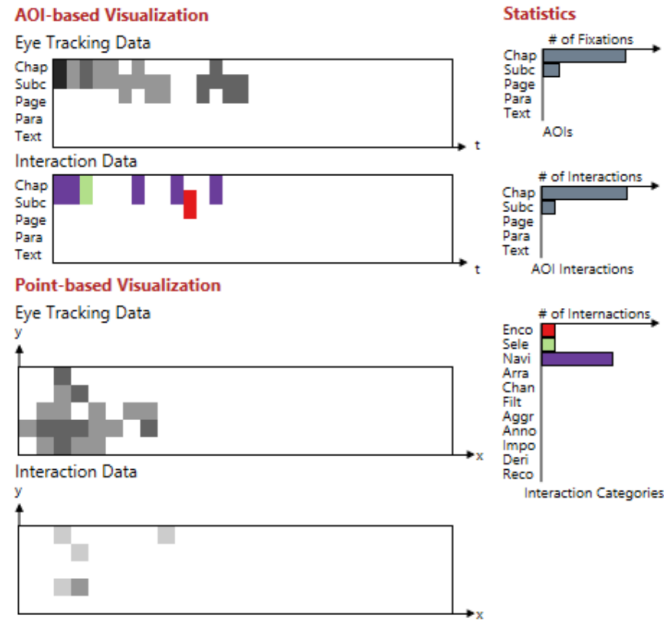


图 1: Data-rich User Behavior

VAST 16, TVCG *TextFile: An Interactive Visualization Tool for Seamless Exploratory Analysis of Structured Data and Unstructured Text* 文章的大话题不错, 但是实际操作上感觉没有那么直击要害, 可能是系统后来针对的问题比较特殊和具体. 本身是希望将结构化的数据和非结构化的文本一起探索, 但其实将文本进行NLP出来以后也几乎处于结构化状态, 主要还是靠人的认知来进行下一步探索. 其文章中提出的模型概览具有一定普适性. 系统比较美观, 并且在网上展现开源.

VAST 16, TVCG *TopicLens: Efficient Multi-Level Visual Topic Exploration of Large-Scale Document Collections* 文章思想很简单, 以至于在美国听报告时大家觉得这个系统两天就能做完. 其实一方面这个想法虽然简单别人确实也没做过, 就是放大镜一样将一个话题里面一部分文章“放大”再聚类; 另一方面, 聚类需要时间, 为了做到responsive或者实时性作者们还是提出了一些改进的算法来实现这个目标. 这两点我认为确实不错, 除此以外诸如类别太多颜色如何分配等问题作者们也确实需要考虑.

在美国认识的华人有吴彤霜(华盛顿大学Jeff Heer学生博一, 原港科大), 陈晴(港科大高年级博士生), Po-Ming Law(佐治亚理工, 原港科大). 都是港科大系列, 之前也读过他们一些文章, Wu今年二作的NameClarifier则是组会刚由我报告过. 他们的英语口语都非常不错. 咨询了他们完成一个项目花的时间, 也一点不比我们的少. 研究的问题小而精, 不吹牛, 方法朴素而有效, 加了微信和邮件以后也会向他们多多请教诸如行为分析或者glyph 设计等我认为他们比较强悍的地方.

在poster展览时与三位外国友人进行了深切交谈, 拍照留念, 年龄大概与我们相仿. 第一位来自瑞典Linköping大学做的是三位平行坐标一个用户实验, 平行坐标在今年不少的论文上出现过, 而其导师也是去年在VIS/TVCG 上发表了一篇综述性质文章, 概括并且展望了PCP的一些特性, 我觉得在后面系统中可以作为某些小的创新点. 第二位来自德国Technischen Universität Darmstadt, 做的是有一个企业投资的项目, 目的是为新闻领域工作者帮助从新闻文本中提取有效信息. 系统开发了将近一年也已得到公司认可, 即将进行内部试用. 从其演示的安然公司邮件的案例来看, 我确实觉得系统简洁, 并且通过对交互和简单NLP方法结合再加可视编码后, 达到了比较显著的效果. 第三位是哈佛大学做的一个帮助人认知的利用

本体论方法的系统, 由于对本体论也不是深入了解, 当时那个小哥也是简单介绍; 但其海报确实做的美观而吸引引人入胜.

后面读论文的计划, 是这样安排的. 首先VAST不少关于行为和时间序列、包含一些预测性分析的文章, 不管是TVCG还是会议都还要过一下. 其次, 开会时infovis我们听的较少, 一些评估综述性文章有带来新的insight; 还有一些关于时序、时空数据、网络或者图设计的文章会看一下.